DOCUMENT RESUME

ED 048 755 56 EM 008 814

AUTHOR	Hsia, Hower J.
TITLE	An Information Theory Approach to Communications and Learning.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO	WP-13
PUB DATE	Apr 69
CONTRACT	OEC-5-10-154
NOTE	30p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Communication (Thought Transfer), *Conceptual Schemes, Information Science, *Information Theory, *Learning Theories, Redundancy

ABSTRACT

Any human communication has intrinsic objectives, to reach maximum communication efficiency and dependability. Because of "equivocation" and "noise," these absolute maximums are unobtainable; only relative maximums are possible. Based on Shannon's information theory principles, various information terms and different forms of "redundancy" are discussed to expound the optimal ratio between input and redundancy which is prerequisite to reaching the relative maximums of efficiency and dependability. The merits and deficiencies of information theory in communication and learning research are examined. The conclusion is that information theory, because of its simplicity and universality, may be one of the best instruments for the development of a general, systematic communication-and-learning theory. Redundancy, in its many forms, is seen as the best device to reduce equivocation and the effects of noise. Information theory differentiates error and equivocation (i.e., error by projection and error by omission) which are of different origins and natures and which may be capable of explaining many communication and learning phenomena such as the relative efficiency of different media. (Author/JK)

ED048755

WORKING PAPER NO. 13

AN INFORMATION THEORY APPROACH TO COMMUNICATIONS AND LEARNING

REPORT FROM THE MEDIA AND CONCEPT LEARNING PROJECT

Wisconsin Research and Development
CENTER FOR COGNITIVE LEARNING

THE UNIVERSITY OF WISCONSIN
Madison, Wisconsin

U.S. Office of Education
Center No. C-03
Contract OE 5-10-154

EM 008814

ED048755

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Working Paper No. 13

AN INFORMATION THEORY APPROACH TO COMMUNICATIONS AND LEARNING

By Hower J. Hsia

Report from the Media and Concept Learning Project
Bruce H. Westley, Principal Investigator

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

April 1969

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Working Paper is from the Media and Concept Learning Project in Program 1. General objectives of the Program are to generate new knowledge about concept learning and cognitive skills, to synthesize existing knowledge, and to develop educational materials suggested by the prior activities. Contributing to these Program objectives, the Media Project, which ended in 1967, was directed toward clarifying the effectiveness of audio, visual, and audiovisual means of transmitting information.

CONTENTS

		Page
	Abstract	vii
I	Introduction	1
	Selection Power in Information Processing	2
	Learning as a Communication Process	3
	Controversy Over Information Theory	4
	Schematic Representation of Information in Communication	5
	Some Operational Definitions for Information Theory Terms	7
II	Redundancy in Communication	9
	Various Meanings of Redundancy	10
	Language Constraint as Redundancy	11
	Between-Channel Redundancy	11
	Memory-Process Redundancy	11
	Principal Functions of Redundancy	12
	Dysfunction of Over- and Under-Redundancy	13
	Other Practical Uses of the Redundancy Concept	13
III	Noise, the Functional Opposite of Redundancy	16
	The Dilemma of Communication	17
	The Relationship of Redundancy to Other Information Properties	18
IV	Conclusion	20
	References	23
Figure 1	Schematic Representation of the Several Quantities of Information Involved in a Communication Situation	6

ABSTRACT

Any human communication has intrinsic objectives, to reach maximum communication efficiency and dependability; the former indicates minimum equivocation (information loss) and the latter signifies minimum error in information transfer. Maximum communication efficiency and dependability are unobtainable because of the existence of equivocation and noise; only a relative maximum can be obtained when all interrelationships among various information properties and redundancy have been precisely known. Based upon Shannon's information theory principles, various information terms and redundancy forms are discussed to expound the optimal ratio between input and redundancy which is prerequisite to reaching the relative maximum of communication efficiency and dependability.

The merits and deficiencies of information theory in communication and learning research are also examined, arriving at the conclusion that information theory, because of its simplicity and universality, may be one of the best instruments for the development of a general, systematic communication-and-learning theory. Emphasis is placed upon both the information and redundancy concepts.

Apart from the explication of information properties, stress is on the inevitability of the occurrence of equivocation and noise in any information processing. To reduce equivocation and the effects of noise, redundancy, in its many forms, is seen as the best device. Information theory differentiates error and equivocation (i.e., error by projection and error by omission) which are of different origins and natures and may be capable of explaining many communication and learning phenomena such as the relative effectiveness of different media. By starting with information theory it is hoped that a general communication theory may emerge.

I

INTRODUCTION

Man is immersed in a sea of information; this is why, out of necessity, "man has invented dramatically effective codes for handling information [Gerard and Duyff, 1964, p. 4]." Effective coding and use of information require systematic rules; rules imply constraint. Wherever there is constraint upon any phenomenon, event, system, or language—such as the orbit of planets, the color of flowers, the computer system, and any written or spoken language—there are laws. Lawfulness indicates order, particularly in human communication. Constraint is the prerequisite of any law, and "every law is a constraint," which infers predictability (Ashby, 1956, p. 130).

Communication is concerned with "human relatedness" (Ruesch and Bateson, 1951). Human relatedness defines the rules of the communication system subscribed to by members of any given community. Without grammatical rules, for example, language would be chaotic and probably incomprehensible. Rules imply constraint, and every constraint constitutes redundancy; for Ashby (1956) constraint is equivalent to redundancy. Hence, the two major problems of communication are information and redundancy.

The purpose of communication is to process information; redundancy facilitates information processing. The purpose of processing information, for Norbert Wiener (1954), is for an organism to adapt to its environment —a process primarily activated for biological survival. Every bit of new information creates a disturbance within the organism and behavioral change, however slight, is the result of it. No organism, however, is endowed with the capacity or desire to process and absorb all information available to it. Each organism, according to its own needs and wants, selects its information. Indeed, information in most, if not all, of its connotations, seems to rest upon the notion of "selection power" (Cherry, 1957, p. 242).

SELECTION POWER IN INFORMATION PROCESSING

Selection power is determined by the strategy an organism adopts in information processing; the adoption of any strategy is decided to a large extent by the consideration of the cost involved in information processing. Ashby (1956) has given a succinct example in answering a question he poses to himself:

> Suppose that, somewhere in the universe (as visible to the astronomer), there is a unique atom; the selector wants to find it. The visual universe contains about 100,000,000 galaxies, each of which contains 100,000,000,000 suns and their systems; each solar system contains about 300,000 bodies like the earth, and the earth contains 1,000,000,000,000,000,000,000 dust particles, each of which contains about 10,000,000,000,000,000 atoms. He wants to find a particular one [p. 261]!

He can proceed, Ashby suggests, in one of the two following manners: (1) he can examine one atom at a time; even with the fastest computer, he would have to spend a century to find the atom; or (2) he can divide the number of atoms in half each time, and ask the question, "Is the atom in this half?" He can find the exact atom in about four minutes. This is precisely the strength of information theory, but of course, few worldly phenomena are arranged in the best possible dichotomous order within a unidimensional system that can be consistently divided by two.

Not only is the selection power manifest in the encoding and decoding processes but, in most cases, it determines the acceptance or rejection of information. The rarity, exclusiveness, or surprise value of information, all of which can be expressed in terms of information or uncertainty, determines how information is to be processed. The exclusiveness of information can be seen in a news "scoop" or an invention. Exclusiveness based on the rarity of information has greater surprise value. Selective power is dependent upon exclusiveness and prediction power (Gabor, 1951). These are actually two sides of a coin: exclusiveness indicates uncertainty and prediction implies redundancy.

Many principles propounded in information theory (Shannon, 1949) with only slight modification seem eminently suitable for the formulation of a general communication theory. Information theory, which has had world-wide impact in many diverse disciplines (Dahling, 1962), is derived from several branches of science, notably cybernetics, a synthesized theory of society—human, animal, and machine (Wiener, 1948). It may also prove to be a means of solving many fundamental problems in human communications and learning.

The evolution of information theory in psychology, biology, neurology, medicine, and engineering, among others, has produced many insights into behavioral problems. At the same time some misconceptions have grown up concerning the use of information theory. New terminology flourishes; in the field of psychology alone, Quastler (1955) made an attempt to unify the nomenclature, and MacKay (1952) also tried to achieve some agreement on the use of information theory terms.

LEARNING AS A COMMUNICATION PROCESS

From a cybernetist's point of view, learning is the reduction of entropy or uncertainty. Norbert Wiener (1954) defines learning as a continuous process, seeking the reduction of entropy:

> Learning is the modification of a behavior pattern on the basis of experience so as to achieve specific antientropic ends. In these higher forms of communication organisms the environment, considered as the past experience of the individual, can modify the pattern of behavior into one which in some sense or other will deal more effectively with the future environment. In other words, the organism is not like the clockwork nomad of Leibnitz with its pre-established harmony with the universe, but actually seeks a new equilibrium with the universe and its future contingencies [p. 48]....The physiological condition for memory and hence for learning seems to be a certain continuity of organization, which allows the alterations produced by outer sense impressions to be retained as more or less permanent changes of structure or function [p. 55].

Since "no single universally accepted 'basic' learning theory exists" (Hunt, 1962, p. 47), learning may be termed the acquisition, retention, and transfer of information, all involved in an antientropic process. Acquisition,

> the primary phenomenon of learning, is a progressive, incremental change in the proficiency of performance by an organism. The direction, rate, and extent of change in the proficiency of performance are functions for continuous presentation of the conditions under which measurement of the change in performance is made [Brogden, 1951, p. 569].

Learning, put more simply, is "the change in performance, associated with practice and not explicable on the basis of fatigue, of artifacts or measurement, or of receptor and effector changes [Hovland, 1951, p. 613]." Or, learning is the process by which an activity originates or is changed through reacting to an encountered situation, provided that the characteristics of the change in activity cannot be explained on the basis of native response

tendencies, maturation, or temporary state of the organism (e.g., fatigue, drugs, etc.) (Hilgard, 1956, pp. 2-8). Thus for psychologists, learning is concerned with how behavior patterns are established and changed; for students of communications, all changes are the results of information processing, and learning is a process of receiving, storing, and transferring information embodied in any communication, processed by any modality or modalities, with a resultant modification of existing behavior. Hunt (1962) specifies concept learning as (a) any situation that involves the "acquisition or utilization, or both, of a common response to dissimilar stimuli", (b) concept formation, (c) concept selection, and (d) concept identification. All of these may be termed as outcomes of information processing.

CONTROVERSY OVER INFORMATION THEORY

MacKay (1952) points out that general information is concerned with the problem of measuring changes in knowledge, and communication is the activity of replicating representations.

> Amount of selective information is evidently a measure of the statistical rarity of a representation and has no direct logical connection with its form of content, except in cases where they affect its statistical status. One word which was unexpected could yield more selective information to a receiver than a whole paragraph which he knew he would receive.

Precisely because information theory is based upon the statistical rarity of elements in any sign, symbol, or signal system, it is both versatile and handicapped: the versatility of information theory in communication and other research has been clearly demonstrated; its handicaps include tedious computation and the complete separation of information from the idea of meaning or significance. Attneave (1959, p. 88) has warned that information theory is "useful in the study of the organism's information-handling processes, [but] other techniques may often be more useful and more appropriate." Mandelbrot (1965) has commented on its inconvenience.

But the usefulness of the theory to communications research lies in the possibility it affords of producing a general, systematized communication theory dealing with information processing. It is not, as G. A. Miller (1953, p. 119) points out, a panacea: the deficiencies in its practical application are fairly well known. Cronbach (1955) seems to offer more despair than hope:

> The power of information theory to stimulate scientific thought, and psychological thought in particular, is simply demonstrated by the number of projects and publications which Shannon's work has inspired....Many studies begin with insight obtained by looking on a certain psychological process as a communication system. These analogies from the information model lead very often to significant generalizations. The language of information theory has been suggestive in our studies of theory of mental tests, for example, but when we attempted to formulate our generalization in the Shannon mathematics rather than simply as analogies, we ran into difficulties. Some results were unreasonable. Others were reasonable, but we could not defend them as more correct than alternative results derived from other mathematical formulation [p. 14].

Notwithstanding all these limitations, the application of information theory to communication research is primarily based upon the following assertions: First, information theory offers the best possibility of formulating a general, systematized communication model and theory, universally applicable. Second, by employing information theory it is possible to determine an optimal information level for any given individual, or any system under a given condition and at a given time; such knowledge can make it possible to curtail the effort, cost, and time involved in information processing.

SCHEMATIC REPRESENTATION OF INFORMATION IN COMMUNICATION

The schematic representation of information and its function structured by Attneave (1959, p. 56) and modified by G. A. Miller (1964, pp. 176–178) may be adapted to serve as a model. For our purposes, the notation system has been slightly changed, to use the terms of Brillouin (1956) rather than those of Shannon (1949); i.e., $\underline{I}$ (in) = $\underline{H}$ ($\underline{x}$), $\underline{I}$ (out), = $\underline{H}$ ($\underline{y}$), etc.

In a communication situation (see Fig. 1),

In = Information transmitted by the communicator (teacher) as input information for the receiver (student).

Out = Receiver's (student's) output information.

I (in) = The average amount of information transmitted by the communicator (teacher) as the average amount of input information for the receiver (student).

I (out) = The average amount of the receiver's output information.

I_{out} (in) = Information transmitted by the communicator but omitted in the receiver's output information, or not recalled by the receiver; it is information lost in transmission (also called equivocation).

I_{in} (out) = Information contained in the receiver's output information but not transmitted by the communicator; it is added, unwanted information, also called "ambiguity", "noise", or "error".

I (in, out) = The total amount of information.

L = The average amount of information correctly recalled in output information or learned by the receiver (student); it is the shared information between I (in) and I (out), i.e., L = I (in) + I (out) - I (in, out).

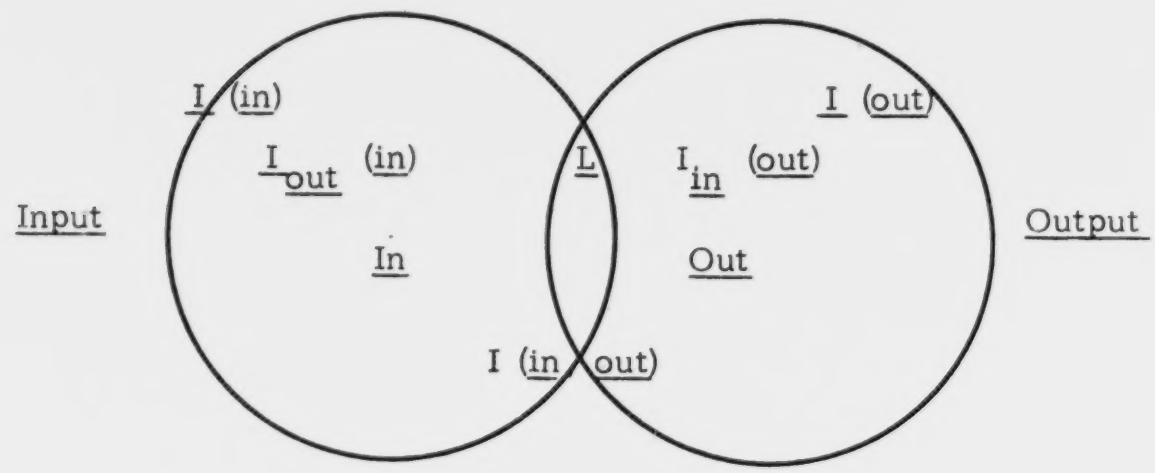

Fig. 1 Schematic representation of the several quantities of information that are involved in a communication situation. After Attneave (1959) and G. A. Miller (1964).

This interpretation of the basic concepts of information theory, as G. A. Miller (1964) points out, is ordinarily used with the object of computing L. The ideal situation is L = I (in) = I (out), which means the student learns everything that is taught and nothing that is not taught. The receiver correctly recalls all the information that is transmitted, indicating maximum learning or perfect information transfer.

The realization of such maximum learning or perfect communication is possible only in theory. First, there is a limit to the amount of information a channel can carry. This limit is called channel capacity, or C.[1]

---

[1] The capacity theorem is given by Shannon (1949) for the discrete noiseless case (p. 28), the discrete channel with noise (pp. 35-39), and for the continuous channel (pp. 64-73).

As I (in) increases, it is possible for I (out) and L to approach the limit C, which is defined as the maximum value that can be reached by I (in, out). In any communication situation, the fundamental objective is to achieve equality; namely, making C = L = I (in) = I (out), indicating maximum utilization of capacity. Because of the existence of equivocation I_{out} (in) and noise I_{in} (out), it is inevitable that C ≥ I (out) ≥ L in a real communication situation. Second, the fallibility of man's encoding and decoding ability requires redundancy in every communication system; therefore, C cannot be utilized to its fullest extent, because redundancy increases at the expense of information. Third, there is no way of completely eliminating equivocation and ambiguity under any rigidly controlled conditions; therefore, it is generally the case that C > L. This further limits L.

SOME OPERATIONAL DEFINITIONS FOR INFORMATION THEORY TERMS

The concepts of information (entropy), noise, equivocation, capacity, and redundancy have usually been expressed in terms of the probable occurrence of elements in a sign system.[2] The amount.of information is usually expressed in terms of "bit," which is the abbreviation for binary digit; a bit is defined as the unit of choice between two equiprobable alternatives. The amount of information or entropy is also synonymous with uncertainty. Information is usually measured in the basic formula:

$$H = -\Sigma p_i \log_2 p_i \qquad (1)$$

The simplicity of this formula has resulted in information theory being termed a "one-theorem theory" (Chung, 1962). The ratio of the actual entropy to the hypothetically possible maximum entropy is called the relative entropy, or relative information. One word of caution is in order: Some information theorists use the term "negentropy" in place of "entropy" (information or uncertainty). For example, "as information about a system increases, uncertainty about it decreases, and so information is the negative of uncertainty. When uncertainty about it is 0, no further information

---

[2] It was Shannon (1949) who pointed out that the mathematical form of the definition of uncertainty or information is identical to the formula for entropy already well known in mechanics (p. 20). Ever since, it has been customary to use the symbol for entropy (H) to stand for information or uncertainty and to use the terms information and entropy interchangeably. This practice is followed in the present paper.

can be received about it [Pask, 1961, pp. 26-27]." This is obviously more of a semantic than a conceptual difference.

Because all information properties—input, output, recalled information, equivocation, and error—can be calculated, the information theory model permits precise determination of amount of information in some situations. But the computation cost for the conditional information of even trigrams, e.g., the dependency of one element upon the two elements preceeding it in a continuous communication, is prohibitive. Thus, information theory may be more fruitfully exploited by adapting its basic concepts and fundamental principles than by its time-consuming calculation.

II

REDUNDANCY IN COMMUNICATION

Apart from information (entropy), the most important concept is redundancy. Redundancy has many interpretations. For such classical information theorists as Shannon (1949), Fano (1949) and Cherry (1957), redundancy is a mathematical concept referring to the proportion of relative entropy contained in a communication. For others, redundancy "implies an overdetermination of meaning [Meier, 1962, p. 123]" and "the repetition of a signal [Smith, 1966, p. 365]"—the latter probably being an oversimplification.

Redundancy in information theory is conventionally defined as one minus the relative entropy as customarily given in the following formula:

$$\text{Redundancy} = 1-(H_{actual} - H_{maximum}) \qquad (2)$$

where relative entropy is the ratio of the actual to the hypothetical maximum entropy, which is derived from the equally probable occurrence of each event, symbol, or unit within a given set. It may be recalled that H_{actual} is the fundamental concept of information theory, i.e., $\underline{H} = -\Sigma \underline{P}_i \log_2 \underline{P}_i$ whereas $H_{max} = \log_2 \underline{N}$. Weaver (1949) says redundancy generally

> is the fraction of the structure of the message which is determined not by the free choice of the sender, but rather by the accepted statistical rules governing the use of the symbols in question. It is sensibly called redundancy, for this fraction of the message is redundant in something close to the ordinary sense; that is to say, this fraction of the message is unnecessary (hence repetitive or redundant) in the sense that if it were missing the message would still be essentially complete or at least could be completed [Weaver, 1949, p. 104].

It would be ideal if redundancy could be eliminated and if information processing with any physical channel could be maximized to the limit of capacity, thereby minimizing the effort and cost involved in information processing. This idea is, however, seldom realized, as evidenced in the

occurrence of error and equivocation. A redundancy-free system in reality is possible only if the entire channel system is completely error-free and has an unlimited capacity. A redundancy-free system is entirely feasible and can be worked out precisely in any physical communication system by devising the most efficient code. A series of error-free systems is theoretically possible but practically unrealizable, as exemplified in the error-detecting systems built into the more sophisticated computers. The checking system in a computer (for example, the use of a binary digit to govern six binary digits and check the sum to determine whether an error has been made) is a redundant system. Due to the noise-congested channel system and the error-prone nervous system of man, redundancy in fact proves to be a boon to any communication system connected with a living organism. A computer instruction code may be designed to be redundancy-free, but a telephone conversation is by no means completely error-free. The implication is that there must be redundancy in any human communication system.

VARIOUS MEANINGS OF REDUNDANCY

> Redundancy is a property of languages, codes, and sign systems which arises from a superfluity of rules, and which facilitates communication in spite of all the factors of uncertainty acting against it . . . redundancy is built into the structural forms of different languages in diverse ways [Cherry, 1957, pp. 18-19 and 118].

Redundancy also exists in a variety of forms in connection with a sign or symbol system processed by one or more modalities. But no study has yet attempted to classify all forms of redundancy; the need for clarification and rigorous definition of the redundancy concept seems to be a matter of primary importance, for not all studies employ identical concepts of redundancy.

For example, as Attneave (1959) has attested,

> Bricker's usage of 'redundancy' is somewhat different from mine. I consider 'redundant' the information which various stimulus components share with one another, whereas for Bricker, information in excess of that necessary for the determination of certain specified naming or categorizing response is 'redundant' [p. 85].

Binder and Wolin (1964) have clarified some misconceptions about redundancy. In the discussion of inconsistency with respect to Shannon's sequential association, Garner (1962) equates redundancy with contingent uncertainty in terms of distributional and correlational constraint. Binder and Wolin (1964), in presenting their concepts of univariate, bivariate, multivariate, and Markovian redundancy, seem to be of the opinion that discrepancy in the redundancy concept arises from the models individual researchers have adopted.

LANGUAGE CONSTRAINT AS REDUNDANCY

The simplest redundancy is in the form of repetition, which is simply, as Cherry (1957, p. 185) says, an addition. This rudimentary principle is also evident in any other form of redundancy. For example, "taxicab" is a redundant representation of "taxi" and "cab." Another major identifiable source of redundancy is the redundancy in a language, for example, English, which is said to be 50 percent redundant (Weaver, 1949, p. 104). Redundancy in language may be termed "constraint" (McCormick, Travers, Van Mondfrans, and Williams, 1964) in a common sense definition, or in a more elegant form, it may be termed sequential redundancy (Shannon, 1949; Garner, 1962).

Constraint may be empirically determined, and sequential redundancy may be mathematically derived. Essentially they are synonyms. Adopting the simpler concept of constraint, a large amount of information can simply be classified as constraint or nonconstraint, without involving calculation of monogram, digram, trigram and so forth to obtain the precise amount of sequential entropy and redundancy. From a practical point of view, obtaining sequential redundancy from a large ensemble of information is prohibitively time-consuming, if not entirely impossible.

BETWEEN-CHANNEL REDUNDANCY

Audiovisual simultaneous presentation raises a theoretical dispute on the interfering and facilitating effects of one channel on the other. But in any dual channel presentation, there is also information shared by both channels; hence, there is also between-channel redundancy. Between-channel redundancy refers here to the redundancy of information between auditory and visual channels. All other redundancy previously mentioned is, in a sense, within-channel redundancy, or content-redundancy. Between-channel redundancy is a discrete source of redundancy, independent of its other forms.

Between-channel redundancy refers to the similarity of information between two channels; it is a bivariate model. Conceivably, between-channel redundancy is unity when both visual and auditory channels transmit identical information; it is zero when the visual and auditory channels emit completely different information.

MEMORY-PROCESS REDUNDANCY

The most important form of redundancy is probably the redundancy between information being processed and information within the memory

system. This redundancy may conveniently be termed "memory-process redundancy." M-P redundancy is sometimes called internal redundancy, as Brown (1959) states, "assuming that when a memory trace is established, it usually has some internal redundancy. In other words, the trace is established with more features than are necessary to represent the information which the trace is required to store." This redundancy, Brown argues, takes up memory storage space, but "the trace system (redundancy) can often supply some of the information lost through decay of an individual trace [p. 734]."

It appears to be reasonable to think that an increase in M-P redundancy should increase the comprehensibility of a communication and make it more accessible to memory. Assuming the existence of noise within the nervous system, the protection of information from loss and interference calls for redundancy; but "sensory information has to be transmitted from place to place in the central nervous system and the reduction of redundancy before this is done would enable the number of internal connecting fibres to be reduced [Barlow, 1959]."

PRINCIPAL FUNCTIONS OF REDUNDANCY

In spite of the variety of its forms, redundancy has the same functional relationship with other information properties. Optimal redundancy can increase efficiency in information processing, retention, and transfer because it is capable: (1) of reducing error to a tolerable level in the encoding process and in the decoding process; (2) of reducing the effects of noise, interference, and distortion in both the external channel and internal (physiological) channels; and (3) of facilitating association and discrimination and helping to prevent forgetting.

These effects may be obtained only if redundancy does not reach the point beyond which it gradually reduces information, and finally replaces it altogether, a condition of "over-redundancy." There is theoretical point beyond which additional redundancy impairs communication efficiency, rather than increasing it. In an experiment on spatial signal patterns, this critical point has been derived (Hsu Lien-Tsang, 1963). A considerable part of the information contained in a signal structure, Hsu points out, would become "redundant" in the course of frequent repetition of the required operation in association with a definite signal. Three experiments involving naming numbers and pointing to lights in tasks varying relative stimulus frequencies (Fitts, Peterson and Wolpe, 1963) lend support to Hsu's interpretation. As redundancy increases, average reaction time to the frequent stimulus components increases. The delay in processing time proportionally raises the cost of information processing.

DYSFUNCTIONS OF OVER- AND UNDER-REDUNDANCY

Over-redundancy, defined as redundancy in excess of optimum redundancy, may introduce what is known as "cue conflicts," which affect the subjects' certainty (Bruner, 1957), although over-redundancy decreases information. Whenever information increases beyond the capacity limit, information loss occurs. Further evidence can be seen from Baker and Alluisi's study (1962), in which subjects listened to examples of the four types of auditory figures to be practiced and made identifications of the corresponding visual figures. It was found that response time to redundant figures was greater than to random figures, and response time to more complex figures was greater than to simpler figures.

Generally, appropriate repetition, which is one form of redundancy, decreases uncertainty and facilitates communication so long as the receiver can still gain some information from repetitions he failed to process in his previous attempts; of course, the difference between $\underline{I}$ (in) and $\underline{I}$ (out) becomes increasingly smaller with each repetition. As more and more information is processed and stored in the central nervous system after each repeated presentation, mere repetition can practically exhaust information as M-P redundancy reaches unity; i.e., all the information that has to be processed is already in the memory. By definition, any communication containing no information at all is considered noise; for example, a parrot capable of speaking only a few words would be very much appreciated by someone who heard him for the first time, but not for the hundredth time.

The English language is so redundant that "a fraction of letters can be randomly deleted from a reasonably long message without making the message unintelligible, 'FOR EXMPLE WENTYIVE PRCET OF HE LTTERS I TIS SENTENCE HVE BEN DT RANM' [Rapoport, 1965]." English prose is still intelligible when all vowels are eliminated, but when the difficulty level, or the amount of information, is increased, as in this deleted sentence: "probblty s mthmtcl dscpln wth ms kn t ths, for xmpl, i gmtry r nlytcl mchncs." This is an example of "under-redundancy." An extreme case of under-redundancy, for instance, is a table of random numbers, which is actually a pure case of zero redundancy. It may be seen that an optimal range of redundancy must be reached to insure intelligibility.

OTHER PRACTICAL USES OF THE REDUNDANCY CONCEPT

The selection of letters in a text is the simplest application of information theory in deriving the redundancy rate of letters, in contrast to

Shannon's "guessing game," in which the subject is to guess the next letter until the right letter is identified. Many studies have been carried out along this general line. It is readily demonstrated that both the "guessing game" and "replacement" with deleted letters involve constraint or redundancy.

Results of many experiments may be fitted into the redundancy concept: the more frequently used words have higher thresholds and are therefore more easily processed (Solomon and Howes, 1951); this is true even in the presence of noise. Reading rates are faster with more familiar words (Pierce and Karlin, 1957), and the number of repetitions required for correct identification is also affected by word frequency (Postman and Rosenzweig, 1957). These results implicitly demonstrate the redundancy principle in different forms.

Experiments which use the redundancy principle without calling it that are numerous; for example, G. A. Miller, Bruner, and Postman (1954) show that the perception of letter sequences is affected by the degree of their resemblance to ordinary English.

Another implicit form of redundancy is readability. Klare (1963) has compiled nearly 40 readability formulas and discussed their features. All these readability formulas are derived from variations of redundancy or constraint and could be summarized in the basic redundancy formula, $\underline{R} = 1 - \underline{H}_{actual} / \underline{H}_{max}$. Where individual differences such as intelligence or vocabulary are concerned, M-P redundancy is involved and the difference can be expressed in terms of prestored information. The rationale for using redundancy as the standard formula for all readability formulas seems quite legitimate, as readability is concerned mainly with the distribution of syllables, phonemes, letters, words, and/or the ratios of any combination of these four. When redundancy is used to replace readability formulas, not only distributional but also sequential redundancy can be known. Thus, a more precise measure of readability can be made available—one which would not be limited only to a particular language.

Recent studies, departing from counting words and frequencies of the occurrence of input information, have shown a high degree of sophistication in the experimental manipulation of redundancy. For instance, Bricker (1955) arranged patterns of lights in a row of five pairs, and administered three conditions of varying amounts of information and redundancy. Deese (1956) gave statistical support to his general substantiation of Bricker's findings that the more redundant figures require greater response time. Attneave (1955) generated a series of symmetrical (redundant) and random patterns by placing dots in a two-dimensional matrix. In their discrimination studies, Fitts, Weinstein, Rappaport, Anderson, and Leonard (1956) and Fitts and Leonard (1957) used various methods of constructing redundant visual patterns, basically of random and constrained forms.

The redundancy of information processed in both the auditory and visual modalities is easy to conceptualize but hard to compute or define. Most studies on audiovisual redundancy (between-channel redundancy) depart radically from the information theory model and avoid assessing the rate of redundancy between channels. Many experiments used the word "redundancy" in a common-sense way and interpret the "redundant relation" intuitively.

III

NOISE, THE FUNCTIONAL OPPOSITE OF REDUNDANCY

Noise introduces uncertainty; in other words, it changes the rate of probability from the more probable, more certain to the less probable, less certain. "If noise is introduced, then the received message contains certain distortions, certain errors, certain extraneous material, that would certainly lead one to say that the received message exhibits, because of the effects of the noise, an increased uncertainty [Weaver, 1949, p. 109].

This is, indeed, a curious aspect of noise, for there is no intrinsic way to distinguish it from information; therefore, it is "purely relative to some given recipient who must say which information he wants to ignore," and "only when some recipient is given...,is distinction between message and noise possible [Ashby, 1956, p. 186]." Noise may be defined more generally without involving subjective judgment: It includes all "static, distortions, disturbances, perturbations, and errors introduced in the signal [Cohen, 1962]." Generally, the presence of noise in information processing reduces output [I (out)] and recalled information (R), and increases error $\left[I_{in} \text{ (out)} \right]$ and equivocation $\left[I_{out} \text{ (in)} \right]$; thus both communication efficiency and dependability are reduced.

By treating noise as a purely physical quantity, the concept of noise may be more clearly illustrated. Based upon the entropy of a Gaussian distribution, Bell (1962) introduces the concept of the "entropy power" of noise, "to take account of nonrandom characteristics" of noise (p. 65). That noise is entropy is well established, but the idea of noise being "undesirable information" may be a misleading value judgment. Noise in our working definition is the fraction of information not sent. The jamming device in radio broadcasting may serve as an illustrative example: the noise generated by the jamming device is signal not transmitted by the communicator.

Another characteristic of noise is that it is capable of changing the elements of a sign or signal system. According to Shannon's examination of the problem:

> This means that the received signal is not necessarily the same as that sent out by the transmitter. Two cases may be distinguished. If a particular transmitted signal always produces the same receivec signal, i.e., the received signal is a definite function of the transmitted signal, then the effect may be called distortion. If this function has an inverse—no transmitted signals producing the same received signal—distortion may be corrected at least in principle. . . [Then] noise is considered to be a chance variable [Shannon, 1949, p. 34].

From either the sender's (teacher's) or receiver's (student's) point of view, noise is "negative information" because "the noise source thus increases the receiver's doubt, and we may regard it as possessing a certain rate of destruction of information ('negative information') [Cherry, 1957, p. 199]."

Noise occurs in any communication. Shannon has suggested possible ways of combating noise by proper coding. "It is possible to send information at the rate C through the channel [with noise] with as small a frequency of errors or equivocation as desired by proper coding [Shannon, 1949, p. 39]." Redundancy which incorporates the concept of proper encoding is probably the best means available to cope with the effects of noise. "Depending upon the type of noise, and type of channel, redundancy is best added in different ways; but the whole subject is very difficult [Cherry, 1957, p. 185]."

On the other hand, noise, the functional opposite of redundancy, does share some of its features. Both decrease communication efficiency and raise the cost of information processing, because both occupy message space and time, forcing the limited capacity to accommodate them. However, their fundamental differences lie in the fact that redundancy reduces equivocation and error whereas noise increases both. To reduce error, one of the manifest results of noise, redundancy is the price any communicator has to pay. However, when information input is low, noise has the unintended function of alerting subjects to the information processing task, thereby facilitating information processing (Samual, 1964; Broadbent, 1964).

THE DILEMMA OF COMMUNICATION

In any communication system, it is generally desirable that input information [I (in)], output information [I (out)], and recalled information (L) be as high as possible, and that error $\left[{}_{\text{in}}I \text{ (out)} \right]$, equivocation $\left[{}_{\text{out}}I \text{ (in)} \right]$, and redundancy (R) be as low as possible so as to minimize cost in terms of information processing time and effort, and to maximize efficiency and dependability. This is clearly the dilemma of human

information processing. Because of the imperfection and fallibility of any human information processing system, equivocation increases linearly with input. The reduction of equivocation necessitates the proportional reduction of input information; the elimination of error or the curtailment of its effects requires various forms of redundancy which, as a rule, decrease input, output, and recalled information and raise the cost of information processing. As input information increases, output approaches the maximum of information processing capacity. When input reaches and increases beyond maximum capacity, output information declines and recalled information levels off.

The remedies for the adverse effects of continuously increasing information appear to be: (1) to keep input information at the optimal rate for the receiver to process; (2) to increase and manipulate redundancy to compensate for the anticipated loss of information and to combat the introduction of noise and the projection of error.

If the rate of efficiency (E) and dependability (D) is increased, it should be the result of reduction or elimination of error and equivocation. The higher rate of E and D serves to economize the cost of information processing. Substantial savings in cost can be realized only when the amount of information and redundancy of input information can be ascertained in advance according to the receiver's capacity and environmental factors, so as to produce an input with optimal redundancy and entropy, thus keeping noise and equivocation at a minimum.

THE RELATIONSHIP OF REDUNDANCY TO OTHER INFORMATION PROPERTIES

As redundancy always increases at the expense of entropy, both entropy and redundancy are required to reach an optimum rate and maintain an optimum ratio if communication is to have maximum efficiency. The rationale for this seems obvious: as redundancy increases, it takes away more and more space which information might otherwise occupy.

Redundancy exists in every language to facilitate information transfer. Regardless of what form it may take, it exists usually at the expense of information and is, in theory, always inversely related to entropy. Its increase invariably brings about a decrease in entropy, and raises the cost of information processing. On the other hand, the increase in redundancy and the correspondent decrease in entropy improves efficiency. The increase in redundancy also curtails the amount of information to be processed, reduces the effects of noise and equivocation, and improves dependability.

Having examined the relations between redundancy and input and between output and input, relations of redundancy to other information properties, notably error, equivocation, and recalled information, become clear. From the fundamental formula for redundancy, it is obvious that redundancy would be unity only when the relative information is zero: actual input information must be zero; consequently equivocation must also be zero. Equivocation maintains a positively linear relation with input beyond the capacity limit; hence it must also be true that equivocation maintains an inverse relation with redundancy when maximum theoretical input remains the same.

The relation of redundancy to $\underline{I}$ (out), error$\left[\underline{I}_{\underline{in}}\ (\underline{out})\right]$, and $\underline{L}$ is subject to numerous contaminating factors, chiefly memory-process redundancy, and has not been determined empirically. However, it generally maintains a logarithmic relation with recalled information, error, and output. As the relations between output and input information and between input and redundancy are known, the formulation of the relation of redundancy to other information properties can be worked out. Relative information, the ratio of the actual input to the theoretically maximum input, determines redundancy. Only variation of the actual $\underline{I}$ (in) in a given communication varies redundancy. Due to the capacity limit and the relation between $\underline{I}$ (in) and $\underline{I}$ (out), the relation between $\underline{I}$ (out) and R is logarithmic. $\underline{I}$ (out) and $\underline{L}$ are positively correlated; therefore, the relation of $\underline{L}$ to redundancy is similar to that between $\underline{I}$ (out) and $\underline{R}$. To a lesser degree, this relation is true also between $\underline{R}$ and $\underline{I}_{\underline{in}}$ (out).

The relationship between redundancy and cost is an uncomplicated one. Other things being equal, the actual information processing rate decreases as redundancy increases; therefore cost increases proportionally with redundancy. On the other hand, input information with no redundancy renders information processing extremely difficult, if not impossible. Hence there must exist an optimum relationship between cost and redundancy, which can be found in the optimum ratio between input information and redundancy.

IV

CONCLUSION

The fundamental problem in communication and learning is to achieve maximum communication efficiency and dependability. Perfect information transmission and reception (perfect communication) is only theoretically possible; by perfect communication is meant that in which the information being transmitted, processed, and fed back sustains no equivocation or error.

The rationale for employing information theory principles in basic communication and learning research seems to be justifiable on these grounds: first, the basic concepts of information theory are applicable to the quantitative aspect of communication and learning. Yet it is also highly general. Wiener (1948) regards all systems as information processing mechanisms. History, culture, and civilization may be thought of as societal information processing. An excellent indicator of cultural achievement, for example, is a person's vocabulary (Meier, 1962, p. 116), and the expansion of vocabulary appears to obey information processing principles precisely. Second, from a universalistic point of view, information theory is not restricted to any particular language; it deals with basic communication elements rather than the structure of particular communication systems; it is adaptable to any sign, signal, or symbol system.

Third, from a physiological point of view, information theory works in a manner similar to the assumed functions of the central nervous system, in which neurons appear to function exactly as information theory predicts (Von Neumann, 1958; Bell, 1962). Fourth, from a pragmatic point of view, only information theory endeavors to quantify information of various kinds in a systematic way; thus, it is capable of dealing with all possible communication functions, including information exchange, storage, and transfer. Fifth, from a scholarly point of view, only information theory, of all other theories in communications, offers a unified, standardized terminology with appropriate measures, rules, and functions. There is reason to hope that a set of communications laws based upon information theory principles may eventually be developed.

Any purposive communication has an intrinsic objective: it is what Weaver calls "the effectiveness problem" (Shannon, 1949, p. 96). Its prerequisite is, however, "the technical problem," to obtain maximum communication efficiency and dependability; the former indicates minimum equivocation (information loss) and the latter signifies minimum error in information transfer from the communicator to receiver. How to arrive at maximum communication efficiency and dependability is a crucial matter in the study of communication and learning. In any information transfer, particularly in human communication, principally because of man's limited capacity in information processing together with other physio-psychological factors, both equivocation and error are inevitable. The inevitability of equivocation and error demands a remedy, and redundancy is probably the most effective means man has found to reduce equivocation and error in information transmission and reception.

The function of redundancy is to curtail equivocation; to reduce error to a tolerable level in both encoding and decoding processes; to lessen the effects of noise, interference, and distortion; to facilitate information association and discrimination; and to reduce forgetting. But the introduction of redundancy into sign systems and channel systems invariably raises the cost of information processing in terms of time and space, as redundancy has to take away message space that information might otherwise occupy. In order to reduce equivocation and error, it is necessary to increase redundancy; but to increase redundancy is to decrease information. This is the dilemma of communication and learning.

The existence of equivocation and error renders perfect communication impossible. By manipulating redundancy in a message and/or between channels, it appears possible to achieve relative maximum information transfer, taking into account entropy, equivocation, error, and redundancy. In other words, information and redundancy must be maintained at an optimal ratio to keep equivocation and error at a relative minimum, and to keep the cost of information processing to a minimum. The maintenance of an optimal ratio between information and redundancy is of fundamental importance in communication and education, for it can increase "selection power." While "selection power" may not be exactly equal to "intellectual power" as Ashby (1956, p. 272) suggests, it may very well be the first step toward achieving maximum communication efficiency and dependability.

REFERENCES

Ashby, W. R. An introduction to cybernetics. New York: John Wiley, 1956 (Science ed., 1963).

Attneave, F. Application of information theory to psychology. New York: Henry Holt, 1959.

Attneave, F. Symmetry, information and memory for patterns. American Journal of Psychology, 1955, 68, 209-222.

Baker, E. J., and Alluisi, E. A. Information handling aspect of visual and auditory form perception. Journal of Engineering Psychology, 1962, 1, 159-179.

Barlow, H. B. Sensory mechanisms: The reduction of redundancy and intelligence. In Mechanisation of thought process, Vol. 2, National Physical Laboratory, Symposium No. 10. London: Her Majesty's Stationery Office, 1959. Pp. 535-574.

Bell, D. A. Intelligent machines. New York: Blaisdell, 1962.

Binder, A., and Wolin, B. R. Information models and their uses. Psychometrika, 1964, 29, 29-54.

Bricker, P. D. The identification of redundant stimulus patterns. Journal of Experimental Psychology, 1955, 49, 73-81.

Brillouin, L. Science and information theory. New York: Academic Press, 1956.

Broadbent, D. E. The approach to information processing through behavior. In Gerard, R. W., and Duyff, J. W. (Eds.), Information processing in the nervous system, Vol. III, Processings of International Union of Physiological Sciences, 22nd International Congress, Leiden, 1962. The Netherlands: Mouton, Co., 1964. Pp. 9-14.

Brogden, W. J. Animal studies of learning. In S. S. Stevens (Ed.), Handbook of Experimental Psychology. New York: John Wiley, 1951. Pp. 568-612.

Brown, J. Information, redundancy and decay of the memory trace. In Mechanisation of thought process, Vol. II, Symposium No. 10. London: Her Majesty's Stationery Office, 1959. Pp. 729-746.

Bruner, J. S. On perceptual readiness. Psychological Review, 1957, 64, 123-152.

Cherry, C. On human communications: A review, a survey, and a criticism. New York: John Wiley, 1957 (Science ed., 1961).

Chung, Kai Lai. The ergodic theorem of information theory. In Machol, R. E., and Gray, P. (Eds.), Recent development in information and decision processes. New York: Macmillan, 1962. Pp. 141-148.

Cohen, J. E. Information theory and music. Behavioral Science, 1962, 7, 137-163.

Cronbach, Lee J. On the non-rational application of information measures in psychology. In Quastler, H. (Ed.), Information theory in psychology. Glencoe, Illinois: The Free Press, 1955. Pp. 14-30.

Dahling, R. L. Shannon's information theory: The spread of an idea. In Katz, E., et al. (Eds.), Studies of innovation and of communication to the public, Studies in the utilization of behavioral science, Vol. II. Stanford U. Institute of Comm. Research, 1962. Pp. 117-140.

Deese, J. Complexity of contour in the recognition of visual form. USAF WADC, TR., 1956, No. 56-60.

Fano, R. M. The transmission of information. M.I.T. Research Lab., Electronics, Tech. Rept., 65, 1949.

Fitts, P. M., Peterson, J. R., and Wolpe, B. Cognitive aspects of information processing: II, Adjustments to stimulus redundancy. Journal of Experimental Psychology, 1963, 65, 423-432.

Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., and Leonard, J. A. Stimulus correlates of visual pattern recognition. Journal of Experimental Psychology, 1956, 51, 1-11.

Gabor, D. Lectures in communication theory. Cambridge, Mass.: Massachusetts Institute of Technology, 1951.

Garner, W. R. Uncertainty and structure as psychological concepts. New York: John Wiley, 1962.

Gerard, R. W., and Duyff, J. W. Information processing in the nervous system, Vol. III, Processings of the International Union of Physiological Sciences, 22nd International Congress, Leiden, 1962. The Netherlands: Mouton & Co., 1964.

Herdan, G. Language as choice and chance. Groningen, The Netherlands: P. Noordhobb, 1956.

Hilgard, E. R. Theories of learning, (2nd ed.) New York: Appleton-Century-Crofts, 1956.

Hovland, C. I. Human learning and retention. In Stevens, S. S. (Ed.), Handbook of experimental psychology. New York: John Wiley, 1951. Pp. 613-689.

Hsu Lien-Tsang. On the influence of redundancy of information in the process of learning spatial signal patterns. ACTA Psychologica Sinica, 1963, 3, 230-238.

Hunt, E. B. Concept learning: An information processing problem. New York: Wiley, 1962.

Klare, C. R. The measurement of readability. Ames, Iowa: Iowa State U., 1963.

MacKay, D. M. In search of basic symbols. In Foerster, H. von, et al. (Eds.), Cybernetics: Circular causal and feedback mechanisms in biological and social systems. New York: Josian M. Macy, Jr., Foundation, 1952. Pp. 181-84.

Mandelbrot, B. Information theory and psycholinguistics. In Wolman, B. B. (Ed.), Scientific psychology: Principles and approaches. New York: Basic Book, 1965. Pp. 550-562.

McCormick, M. C., Travers, R. M. W., Van Mondfrans, A. P., and Williams, F. E. Research and theory related to audiovisual information transmission, Interim Report. Salt Lake City: U. of Utah, Bureau of Educational Research, 1964.

Meier, R. L. A communication theory of urban growth. Cambridge, Mass.: The MIT Press, 1962.

Miller, G. A., Bruner, J. S., and Postman, L. Familiarity of letter sequences and tachistoscopic identification. Journal of General Psychology, 1954, 50, 129-39.

Pask, G. An approach to cybernetics. London: Hutchinson, 1961.

Pierce, J. R., and Karlin, J. E. Reading rate and the information rate to human channel. Bell System Technical Journal, 1957, 497-516.

Postman, L., and Rosenzweig, M. R. Perceptual recognition of words. Journal of Speech Disorders, 1957, 22, 245-253.

Quastler, H. (Ed.) Information theory in psychology: Problem and methods. Glencoe, Ill.: Free Press, 1955.

Rapoport, A. What is information? In Lerbinger, O., and Sullivan, A. J. (Eds.), Information, influence, and communication. New York: Basic Books, 1965. Pp. 225-39.

Ruesch, J., and Bateson, G. Communication: The social matrix of psychiatry. New York: W. W. Norton, 1951.

Samual, W. M. S. Noise and the shifting of attention. Quarterly Journal of Experimental Psychology, 1964, 16, 263-267.

Severin, W. J. Cue summation in multiple channel communication. Unpublished doctoral dissertation. Madison: University of Wisconsin, 1966.

Shannon, C. E. The mathematical theory of communication. In C. E. Shannon and W. Weaver, The mathematical theory of communication. Urbana: University of Illinois Press, 1949. Pp. 3-91.

Smith, A. G. Communication and culture. New York: Holt, Rinehart and Winston, 1966.

Solomon, R. L., and Howes, D. H. Word probability, personal values, and visual duration thresholds. Psychological Review, 1951, 58, 256-70.

Von Neumann, J. The computer and the brain. New Haven: Yale U. Press, 1958.

Weaver, W. Recent contributions to the mathematical theory of communication. In C. E. Shannon and W. Weaver, The mathematical theory of communication. Urbana: University of Illinois Press, 1949. Pp. 95-117.

Wiener, N. The human use of human beings: Cybernetics and society. New York: Doubleday, 1954.

Wiener, N. Cybernetics: Or control and communication in the animal and the machine. New York: John Wiley, 1948.